160-TP-013-001

The Role of Metadata in EOSDIS

Technical Paper--Not intended for formal review or government approval.

March 1997

Prepared Under Contract NAS5-60000

RESPONSIBLE ENGINEER

Tom Dopplick /s/	3/28/97
Tom Dopplick, Technical Editor Science Data Engineering Office EOSDIS Core System Project	Date

SUBMITTED BY

Joe Senftle /s/	3/28/97
Joe Senftle, Manager Science Data Engineering Office EOSDIS Core System Project	Date

Hughes Information Technology Systems
Upper Marlboro, Maryland

Abstract

This Technical Paper was written in preparation for the April 1997 EOSDIS Pre-Launch Metadata Workshop sponsored by NASA. The purpose of the paper is to provide an easy-to-read summary of the process used to model Earth science data in EOSDIS including the use of descriptive data (i.e., metadata) to provide expanded services to end users. A recurring theme from users about the design of the Earth Observing System (EOS) Data and Information System (EOSDIS) is that available documents, that describe Earth science data modeling, have too many acronyms, heavy engineering emphasis, and few summaries suitable for end users. This paper is intended to provide summary information about data modeling for end users not familiar with the detailed technical documents available from ECS.

Keywords: metadata, EOSDIS, ECS, data, model

This page intentionally left blank.

Contents

Abstract

1. Introduction

2. Metadata and Large Data Systems

3. Key EOSDIS Metadata Requirements

4. Design Process for EOSDIS Metadata

5. Operations Concept for EOSDIS Metadata

6. Responsibility for Preparation of EOSDIS Metadata

7. Tools and Processes for Preparing EOSDIS Metadata

8. Maintenance and Updates of EOSDIS Metadata

Figures

Tables

This page intentionally left blank.

1. Introduction

1.1 Purpose

This Technical Paper was written in preparation for the April 1997 EOSDIS Pre-Launch Metadata Workshop sponsored by NASA. A recurring theme from users about the design of the Earth Observing System (EOS) Data and Information System (EOSDIS) is that available documents, that describe Earth science data modeling, have too many acronyms, heavy engineering emphasis, and few summaries suitable for end users. This paper is intended to provide an easy-to-read summary of the process used to model Earth science data in EOSDIS including the use of descriptive data (i.e., metadata) to provide expanded services to end users.

1.2 Organization

This paper is organized into the following broad categories:

Chapter 1. Introduction

Chapter 2. Metadata and Large Data Systems

Chapter 3. Key EOSDIS Metadata Requirements

Chapter 4. Design Process for EOSDIS Metadata

Chapter 5. Operations Concept for EOSDIS Metadata

Chapter 6. Responsibility for Preparation of EOSDIS Metadata

Chapter 7. Tools and Processes for Preparing EOSDIS Metadata

Chapter 8. Maintenance and Updates of EOSDIS Metadata

1.3 Acknowledgments

This paper was conceived and produced through close cooperation between ECS and ESDIS staff. After joint development of the outline, draft input was produced by the following ECS writers: Jeff Burt, Cheryl Croft, Ed Dombrowski, Tom Dopplick, Judy Feldman, Siri Jodha Khalsa, Larry Klein, Bob Plante, and Steve Shantzis. Many people gave of their valuable time to review and provide comments on the paper. ECS Reviewers included Karl Cox, Cheryl Croft, Ed Dombrowski, Judy Feldman, Chuck Gross, Tim Gubbels, Janet Hylton, Larry Klein, Dave Manion, Steve Marley, Bob Plante, Joe Senftle, and Steve Shantzis. ESDIS reviewers included Dan Marinelli, Robin Pfister, H. Ramapriyan, Skip Reber, Darrel Roos, and Stan Scott. Additionally, Paul Bailey graciously gave of his time to also review the paper. Tom Dopplick, as technical editor, was responsible for integrating draft input and review comments into the final document.

1.4 Review and Approval

This Technical Paper is an informal document approved at the Office Manager level. It does not require formal Government review or approval. Questions regarding technical information contained within this Paper should be addressed to Tom Dopplick, 301-925-0333, tom@eos.hitc.com.

Questions concerning distribution or control of this document should be addressed to:

Data Management Office
The ECS Project Office
Hughes Information Technology Systems
1616 McCormick Dr.
Upper Marlboro, MD 20774

2. Metadata and Large Data Systems

A recurring theme from users about the design of the Earth Observing System (EOS) Data and Information System (EOSDIS) is that available documents that describe Earth science data modeling have too many acronyms, heavy engineering emphasis, and few summaries suitable for end users. This paper is intended to provide an easy-to-read summary of the process used to model Earth science data in EOSDIS including the use of descriptive data (i.e., metadata) to provide expanded services to end users. But before plunging into the data modeling process (Section 4) it's useful to examine how data and metadata are used in typical large data systems (this section) and the unique metadata requirements of EOSDIS (Section 3).

Large data systems such as EOSDIS, often are involved in "publishing" millions of pieces of data to many distributed users. For the EOSDIS Core System (ECS), publishing will involve ingest, processing, storage, access, and distribution of hundreds of millions of satellite images and other satellite and *in situ* observations, including documentation, to thousands of distributed users. ECS publishing is primarily concerned with managing production of Earth science data, developing access pathways to the data, and order and delivery of Earth science data to end users.

For simple information systems publishing is easy. In a simple system one inspects the actual data records and selects the few data records of interest. For example, suppose one has 10 vacation pictures taken in the Rocky Mountains. One can shuffle through all 10 pictures in an album and select the one of interest by inspection. For a thousand pictures, inspection of each picture would not be efficient; and with a million pictures it's probably not even possible.

The solution is to organize the pictures, and there are three fundamental reasons for adopting this solution: 1) as the number of pictures increases, one can instruct someone else to aid in the search and selection process, 2) there needs to be a convenient way to aggregate pictures that share common themes, and 3) one needs to remember details about pictures that are removed from the album. In general, for very large data systems, three key mechanisms are needed to implement an organized approach for large scale data "publishing" as shown in Figure 2-1.

Key Mechanisms for Organizing Data
1) Identify data so it can be unambiguously handled by an agent, human or machine
2) Institute a robust data discovery process
3) Maintain context for data which has been delivered to the end user

Figure 2-1. Three Mechanisms for Organizing Data in Large Data Systems

Most large data systems support these three mechanisms through the concept of ‘data about the data’ which has come to be called ‘metadata’. In ECS, each of the three mechanisms for organizing data is associated with a particular type of metadata as discussed below.

2.1 Identifying Data

To perform the identification function, ECS has developed a system for assigning unique identifications to all data items (called ‘granules’) and groups of data items (called ‘collections’). The IDs are used as labels by attaching them to the data, and also reside in a directory. In a system with few holdings, or a system conveniently located to the user, inspecting the data is possible. In a very large system, assigning unique IDs becomes critically important, since distinguishing between data items is highly time consuming - both in inspecting each granule and in differentiating between highly similar but non-identical granules. The ID is the tag by which a specific granule can be fished from the system, and is a key metadata component of any directory system.

In the vacation pictures example, all the pictures could simply be placed in a box, which one could pull out and look through occasionally, paying particular attention to favorite pictures. However, if one were to give an agent instructions - for example over the phone - on locating a particular picture from the box, it would be awkward and time consuming to continually describe the pictures. Instead, one could resort to unique picture labels. “Get the picture with lots of blue sky with my friend standing in front of snow capped mountains” is a lot tougher than “Get picture 34”.

2.2 Data Discovery Process

To perform the discovery function, ECS has identified data characteristics - the technical term is ‘attribute’ - which are used to advertise collections as well as search across and within collections. Each type of Earth science data in the ECS system - i.e., collections, granules, browse, delivered algorithm packages, documents, production histories, and QA information - has its own set of attributes to describe a specific item or group of items. For example, collection attributes include the name of the instruments contained in the collection, the name of the organization which is steward for the collection, the spatial and temporal coverage of the entire collection, and other characteristic attributes. Granule attributes describe characteristics that vary for each granule of the collection including variations in spatial and temporal coverage, quality information about each granule, and other granule characteristics.

These characteristic attributes are cross referenced in tables held in ECS computers. The cross reference can be accessed and searched externally to find the IDs for members of groups of data objects - for example, science granules - which are described by common sets of characteristic attributes. So, for example, a list of all granules from a particular instrument can be easily located in the system by checking the cross reference. Alternatively, if the ID is known, a user can use the cross-reference to abstract summary information - the characteristic attributes - about the data.

Returning to the vacation pictures analogy, one may not want to remember the ID of each favorite picture, so one could classify each picture by themes. Then a cross index can be built. Picture 34 contains mountains, blue sky, snow, and friend. So it is listed under each of the appropriate characteristics. Now one can locate the IDs of all the ‘snow’ pictures. In the future if someone wanted to locate pictures which contain ‘snow’ and ‘mountains’, first they would find the list of snow pictures; then check to see which also contain mountains. A computer can, of course, do this much faster than a person.

If someone else wants to organize the vacation pictures for a different purpose - say to compare the quality of cameras - they might want to create different characteristic attributes such as camera model, lens, available light, etc. Mountains may not be important but snow may be very important because it creates difficult lighting conditions for the camera. So it’s not surprising that different people reach different conclusions about which characteristics or attributes are most important. Attributes are human imposed abstractions, and the interesting attribute set depends on human assessment.

Cross referencing systems require a substantial input of intellect, since the cross referencing activity involves the development and assignment of classifications - a uniquely human endeavor that machine intelligence is only slowly penetrating. Because classification systems represent intellectual property of the highest order, collaborations among a wide group developing classification systems represents a substantial challenge. Moreover, the theoretically desirable cross-referencing system and the practically attainable indexing system generally diverge due to hardware and software limitations - and in a large system they may diverge substantially. The result is the need to select the ‘best’ set of attributes. By working iteratively with the user community, ECS has tried to select the best practically attainable attributes and represent them in the ‘ECS Data Model’.

2.3 Maintaining Data Context

To provide data context, ECS attaches an ID label (physically or electronically) and a full record of the characteristic attributes held in the ECS cross referencing system to data that are distributed to end users. This is practical since it gives the user the convenience of recalling the attributes without the need to re-search ECS. The format which ECS uses to package the data - HDF and HDF-EOS[1] - is organized to allow the metadata to be packed directly with the science data file.

Again, returning to the vacations analogy, it’s good practice to write the picture ID and other characteristics on the back of the picture. That way, if someone requests a copy of the picture, they don’t have to keep checking with the owner for cross reference information. For example, if the requester’s interest is obtaining snow pictures, it may be true that the recipient could simply look at the picture to see if it contains snow. However, they could not easily tell what lens was used. So placing this metadata information on the back of the picture solves the context problem when the picture is sent to a recipient unfamiliar with creation of the picture.

[1] Hierarchical Data Format (HDF); see Figure 5-3 for an example of HDF and HDF-EOS internal components.

This page intentionally left blank.

3. Key EOSDIS Metadata Requirements

EOSDIS metadata must describe the expected breath and diversity of all Earth science data types and other persistent data that will be accessed directly or indirectly by end users. Flexibility, extensibility, and evolvability are important since new data descriptions and relationships will likely emerge as a result of interdisciplinary research.

EOSDIS metadata should be selected and organized to permit rapid search and access across collections, or among the individual granules of a collection, based on common query patterns. Search and access should use descriptive information such as the physical variable measured, the instrument or sensor which gathered the data, and characteristics, such as wavelength bands and spatial and temporal coverage, as well as the type of processing.

Collection descriptions should provide sufficient information to allow a user to understand the broad characteristics of the collection; a user should be able to easily choose between available collections. Within collections, or among related or similar collections, users may desire to retrieve coverage for specific areas or time periods, representing subsets of one or more collections.

Users will assemble descriptions of data granules from one or more collections and use this information to further refine the search and access process. Descriptive granule information needs to include specification of granule temporal and spatial coverage, as well as information on the sensor, instrument, and platform utilized in its gathering, and on data quality.

Beyond simple identification of data, a user may wish to preview samples such as browse data, in order to better assess its usefulness. This would be impractical using individual data granules, since data granules can be very large and often the choice is between many granules. Thus, the presence of lower resolution browse granules provides users rapid access to manageable samples.

To better locate data of interest, understand its origins and intended usage, and to better utilize the data, users may also want to access more detailed descriptive or background information. This could range from documentation about the instruments or techniques used in processing the instrument data, to descriptions of the collection, to research papers published by other users of the data. Detailed descriptions of one or more collections should be provided as guide documents, while published papers and other research documents should be kept as a ‘library’, both fully accessible online.

User searches for data could also require information derived from the content of a collection. Selected measurements or statistics, summarized from one or more data granules, could help users locate data regarding phenomena or features of interest. Metadata describing interrelationships among granules, preselections of granule groupings, or temporal sequential information, would also speed users’ access to data of interest, saving them extensive searching and processing time. These classes of metadata would be stored as inventory characterization and summary statistics metadata.

Further, EOSDIS metadata should provide users with information about the format, storage requirements, cost, if any, distribution options, and points of contact for further information about the data. Since collections can have many granules and large volumes, tools are being developed both within and outside of EOSDIS that will help users manipulate, view, and process data. These tools should be identified within the metadata, providing users guidance in how and where they can access the tools.

Finally, EOSDIS services and data will be distributed across the EOSDIS Distributed Active Archive Centers (DAACs) to take advantage of science and data expertise at NASA, USGS, and University data centers. However, information that describes the distribution must be captured in the metadata to facilitate location of services and data. Also, EOSDIS must be interoperable with other Earth science information systems such as the Global Change Master Directory and the EOSDIS V0 (working prototype at the DAACs). EOSDIS must provide users with gateways into other information systems that also support global change research.

4. Design Process for EOSDIS Metadata

4.1 Overview

The purpose of ECS data model is to try to encompass the expected breath and diversity of Earth science data types and other persistent data that will be accessed directly or indirectly by Earth science users. The ECS Data Model captures Earth science data and metadata requirements, highlighted in Section 3, in a logically structured format. This allows system-wide designers, developers and managers to build ECS with the necessary standardization for total system interoperability within a heterogeneous open system environment. The standard information architecture design process, widely used to organize and describe data and metadata in large data systems, was used to create the ECS Data Model.

The ECS design process followed the classic three schema approach (Figure 4-1) using an object-oriented methodology that is a powerful new tool for designing distributed information systems. However, the representation of the data model in DID311[2] using object-oriented charts and terminology, although technically precise, is certainly not user-friendly or intuitive.

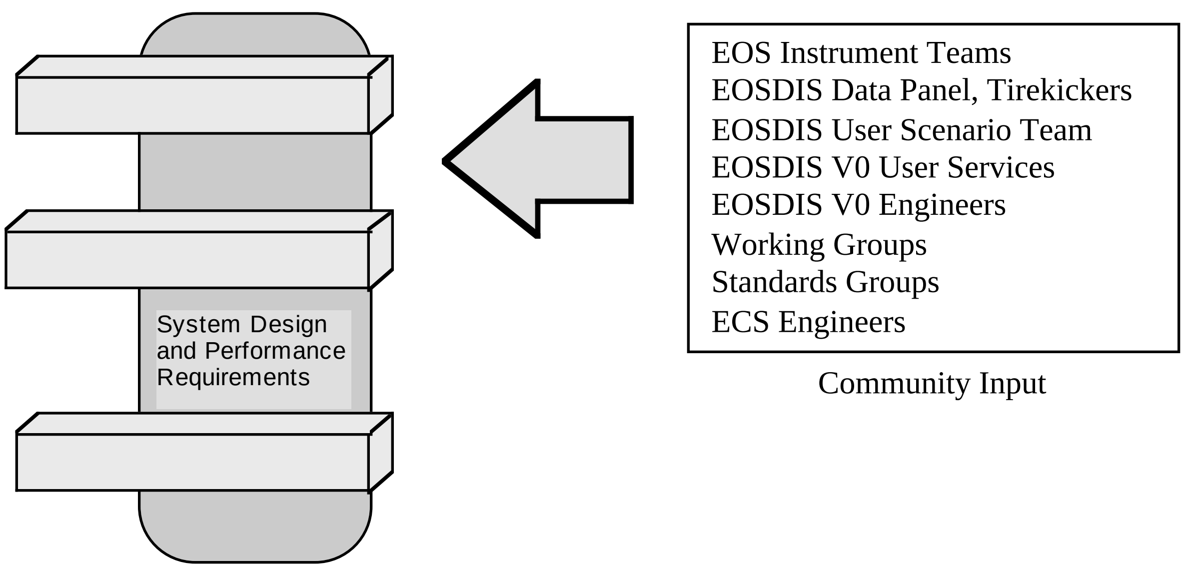

Figure 4-1. Establishment of a Three-Schema ECS Data Model

---

[2] Release B Science Data Processing Segment (SDPS) Database Design and Database Schema Specifications for the ECS Project, 311-CD-008-001, July 1996; ECS documents are available at URL http://edhs1.gsfc.nasa.gov

In Figure 4-1, the *User/External View* is a logical model that represents the end user perception of the data. The *Conceptual Data Model* describes the data assets of the organization illustrating the relationships among the classes of data and specifying the attributes and characteristics of the data. The *Physical Data Model* is the platform dependent representation of the conceptual data model that is implemented in ECS using commercial-off-the-shelf database management systems (DBMS). The ECS strategy is to develop a data model that is capable of describing nearly any type of Earth science information although it is understood that not all attributes would be populated.

Involvement of the user community during the design process began with informal discussions and continued with information exchanges through the Data Model Working Group (DMWG). Requirement recommendations were solicited and iterated with representatives of the EOS instrument teams, interdisciplinary science teams, EOSDIS V0, ESDIS, and other affected organizations. To date ECS has received 232 data model recommendations from sources shown in Figure 4-2; 92% of the recommendations came from non-ECS sources.

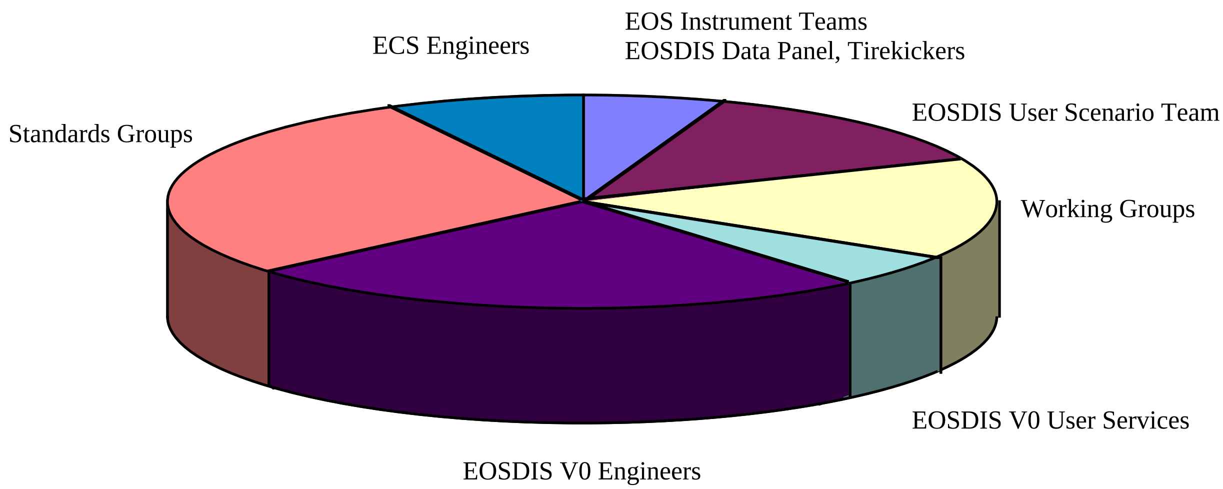

Figure 4-2. Sources of Recommendations for the ECS Data Model

4.2 Application of Data Design Principles to EOSDIS

4.2.1 User/External View

To understand the user view of Earth science data, we began by examining descriptions of Earth science data from NASA, USGS and other Earth science data centers. These data were analyzed with the support of the data centers and scientists associated with the centers. This resulted in a compilation of the universal set of attributes supported by the data centers. This was augmented by coordination with standardization communities such as the Federal Geographic Data Committee (FGDC). Additionally, the Global Change Master Directory (GCMD), Spatial Data Transfer Standard (SDTS), and NASA's Science Data Plan were valuable sources of input to this activity. After careful review and iteration with the science community, metadata objects and associated characteristics were defined that could describe each Earth science data type and other related persistent data expected to be resident in EOSDIS.

4.2.2 Conceptual Data Model

The user view served as the framework for the next phase of ECS Data Model development - incorporating functions and services into the ECS Data Model This activity consisted of mapping the user view to the conceptual data model to ensure that the model provided sufficient functionality to meet EOSDIS system-wide requirements. Where inconsistencies, or omissions occurred, or additional information was required, the metadata objects, attributes, and/or relationships were modified to support the evolving conceptual model. This was an iterative activity which required continual collaboration with the ECS subsystems that were under development, as well as continued involvement with the EOSDIS community. The DMWG provided a forum for information exchange; other technical discussions were held with data producers and user representatives.

The result of this activity is the current conceptual ECS Data Model that describes the metadata objects, metadata attributes and relationships for the Earth science data types, and other persistent data resident in EOSDIS. The metadata objects include collection, granule, browse, delivered algorithm package, documents, production history, and QA information. There are a total of 287 metadata attributes available to describe Earth science data types but only a portion of the attributes are needed to describe a particular data type as discussed in Section 4.3 below. Different subsets of the 287 attributes are used in the descriptions of different data types.

Note that the metadata attributes which the EOS instrument teams are asked to provide are in many cases broader than they provided in the past for their instruments. The reason: additional attributes were added to the data model to support interdisciplinary science research which requires broader characterization of the EOSDIS data holdings. EOS instrument teams have knowledge that must be captured via the data model so that this specialized information can be made known to the broader Earth science community.

4.2.3 Physical Data Model

The physical data model is developed by mapping the conceptual model to the physical schema. Physical data designers used EOSDIS V0 usage statistics to initially design the DBMS tables based on past access patterns. Tuning is performed during ECS development test and will continue throughout operations as actual usage statistics are developed from user interactions with EOSDIS.

4.3 ECS Data Model Vocabulary

The size, richness, and complexity of the EOSDIS data holdings make it imperative to be precise about describing the data and information with sufficient flexibility and extensibility to handle future Earth science data as well as historical data. To fully support ECS publishing requires the development and use of a complete vocabulary that can describe the Earth science data types and other persistent data in EOSDIS. The ECS Data Model vocabulary contains attributes that data providers **may** select to describe relevant characteristics; if a descriptive attribute is not relevant to the data, there is no reason to use that part of the vocabulary. For example, ECS collection metadata for an ocean product allows an entry for depth resolution. If depth resolution is not a property of the data, there would be no entry for depth resolution in the collection metadata. Of course, certain information such as name must be present in all collections in order for the reference system to function correctly.

Because of the large number of EOSDIS users, it is important that the precise meaning of metadata descriptions in the vocabulary be made visible to users. In ECS this is accomplished by using a system-wide data dictionary. The data dictionary provides an online source that users can draw upon when they need clarification or help in understanding the vocabulary. This way users don't have to learn the entire vocabulary before they can start interacting with EOSDIS.

4.3.1 Collection Vocabulary

Collection attributes are used to choose overall characteristics of the member granules that make up the collection. Collection metadata provide a powerful tool to allow quick determination if a collection is relevant to user needs and provides efficient access to the EOSDIS metadata inventories, and ultimately the data in the archives.

General Description Attributes: Includes attributes such as name (e.g., TOA/Cloud Product), discipline (e.g., Earth science), topic (e.g., atmosphere), term (e.g., cloud), variable (e.g., cloud height), and parameter details (e.g., unit of measure), collection version, processing level, comments on quality, and services ECS provides for data that belong to the collection.

Data Origin Attributes: Includes summary attributes that aid in answering the questions how and why data were collected. Attributes include campaign, platform, instrument, sensor, and their descriptions. Data origin attributes describe overall characteristics of the collection; individual members can have varying data origin characteristics, such as varying platforms, and these variations are captured in the granule metadata.

Spatial Coverage Attributes: There are multiple ways to describe spatial coverage of the collection. Attributes are available for bounding shapes of various kinds, map projections, geodetic model parameters, orbit model parameters, orbit numbers, latitude and longitude, radius, bearing, altitude, and others. Only one way is required. Spatial coverage of the collection describes the spatial characteristics of the overall collection, i.e., the spatial coverage envelope for the collection. Individual members can have varying spatial descriptions which are captured in the granule metadata.

Temporal Coverage Attributes: There are multiple ways to describe time coverage of a collection. Attributes are available for a single point in time, a continuous range of time from a start date and time to an end date and time, and a series of times with a start date and time followed by on-and-off cycles with given periodicity. Also available are measures of precision with-respect-to time coordinates. Only one way is required. Temporal coverage of the collection describes the temporal characteristics of the overall collection, i.e., the temporal coverage envelope for the collection. Individual members can have varying temporal descriptions which are captured in the granule metadata.

Contact Attributes: Includes attributes for identifying the responsible organization such as name, title, organization, address, email, phone number, hours of service, and special instructions.

4.3.2 Granule Vocabulary

Granule attributes are used to choose characteristics that vary by granule. There are eight special attributes that allow data producers flexibility to add and define new granule-specific attributes, i.e. Product Specific Attributes (PSAs), as extensions to common-core attributes. Granule metadata, i.e., attributes and their values, allow a user to search through the physical inventory tables in the DBMS that contain the granule metadata. Granule metadata are often called inventory metadata because they ultimately are inserted in the inventory tables. When the data granules of interest are found, additional detailed information can be requested and presented to help scientists better understand the data and locate supporting documentation.

General Description Attributes: Includes attributes such as such as name, size estimate, when produced, process input, ancillary data, associated browse, and various measurements of quality.

Data Origin Attributes: Includes attributes that aid in answering the questions how and why data were collected for each granule including campaign, platform, instrument, sensor, model, and other descriptions.

Spatial Coverage Attributes: There are multiple ways to describe spatial coverage of the granules in a collection. Attributes are available for bounding shapes of various kinds, map projections, geodetic model parameters, orbit model parameters, orbit numbers, latitude and longitude, radius, bearing, altitude, and others. Only one way is required.

Temporal Coverage Attributes: There are multiple ways to describe time coverage of the granules in a collection. Attributes are available for a single point in time and a continuous range of time from a start date and time to an end date and time. Also available are measures of precision with-respect-to time coordinates. Only one way is required.

4.3.3 Browse Vocabulary

Browse attributes are used to describe the browse granule, state the size of the browse granule, and include a pointer to the full resolution data granule.

4.3.4 Delivered Algorithm Package Vocabulary

Delivered Algorithm Package attributes answer the question of how the data were processed in ECS. Attributes include the algorithm name, a pointer to the Algorithm Theoretical Basis Document (ATBD), software version, code, calibration files, configuration information, other documentation, test plans, test data and test results. The Delivered Algorithm Package was included in the ECS Data Model to ensure that other scientists understand how the data were generated and to ensure that scientific results are verifiable and repeatable.

4.3.5 Documents Vocabulary

Document attributes provide pointers to different kinds of documents that are associated with the data. Pointers are available for Algorithm Theoretical Basis Documents, data guides, reference papers, journal articles, analysis reports, and other documents. The document provider chooses which document attributes are applicable to their document. Documents aid users in understanding the meaning and significance of the data.

4.3.6 Production History Vocabulary

Production history attributes provide pointers to reports and logs that are associated with planning and processing EOS instrument data. Selectable attributes point to processing reports (processing status, errors, and production plans) produced by the ECS Planning Subsystem, or to the processing history log (actual time, size, resources used by processing software) produced by the ECS Data Processing Subsystem.

4.3.7 QA Information Vocabulary

QA information attributes exist for both collection and granule metadata. Collection-level QA information attributes include pointers to QA related documents and review material. Granule-level QA information attributes include multiple flags (automatic quality flag, operational quality flag, and science quality flag), QA statistics (percent missing data, percent out of bounds data, percent interpolated data, percent cloud cover), processing QA description, and pointers to browse products and QA granules supplied by data producers. QA granules can be prepared and submitted by EOS instrument teams to add more extensive quality information after inspection of processed EOS instrument data.

4.4 Expected Variability in Preparing Metadata

The number of attributes that must be prepared varies greatly depending on the Earth science data type. Consider the attributes that must be prepared for two data types produced during EOS instrument data processing: interim files and distributable granules. Interim files are used to support EOS instrument processing within ECS and are not useful to end users. Interim files require preparation of ~12 collection attribute values (out of 156) and ~6 granule attributes (out of 66; values to be filled in during processing). An additional 3 link (i.e., pointer) attribute values will be set automatically during science data processing.

Distributable granules are output from EOS instrument processing and typically include preparation of ~100 collection attribute values (out 156), ~38 granule attributes (out of 66; values to be filled in during processing), and 3 browse granule attribute values, if applicable. If documents are provided to the DAAC, the DAAC will prepare 10 document attribute values per document. After science software integration and test, the DAAC will prepare 12 delivered algorithm package attribute values for each processing algorithm that produces a distributable granule. Up to 58 links will be set either automatically or manually at the DAAC to associate collections, granules, documents, delivered algorithm package components, production history, etc.

In EOSDIS, interim EOS files represent the minimum requirement for attributes; whereas, distributable EOS granules represent the maximum requirement for attributes. All other data types in EOSDIS will fall somewhere between these two extremes.

4.5 Validation of the ECS Data Model

The design of the ECS Data Model evolved through a rigorous engineering process with input from both data producer and user representatives. Now validation of the data model becomes an ongoing process that seeks to determine if, in fact, the best practically attainable attributes are represented in the model. Initial validation of the physical model involved comparisons with EOSDIS V0 usage. Scenario-based ECS testing and, later, actual EOSDIS operations will provide direct, measurable assessment of the validity of the model. Another valuable input to validating the model would be an independent assessment by community representatives.

Results from scenario-based testing, EOSDIS operations, and an independent community assessment will provide important inputs needed to make evolutionary changes and refinements to the data model. However, recommendations for design changes (e.g., changing a length or type of attribute value, changing multiplicity of relationships, etc.) must be evaluated by a formal change control process that includes a system-wide impact assessment due to the rippling effect of data model changes throughout ECS as shown in Figure 5-1.

Further, any proposed changes to the ECS Data Model must always be evaluated as to how the proposed change contributes to the goals of the Earth Observing System, NASA’s major contribution to Mission to Planet Earth.

This page intentionally left blank.

5. Operations Concept for EOSDIS Metadata

Before a new Earth science collection can be added to ECS, an Earth Science Data Type (ESDT) descriptor file must be completed and submitted to Science Data Server, a component of the Data Server Subsystem. The ESDT descriptor file is parsed into components and used in various ECS subsystems as shown in Figure 5-1. The ESDT descriptor file includes collection metadata values (for the attributes), granule metadata attributes, valids/ranges for granule metadata attributes, and a list of services for the collection. See Section 6 for discussion of responsibilities for preparation of the collection and granule metadata.

Collection metadata are stored in directory tables within Science Data Server (for distribution with ordered data); stored in directory tables within Data Management Subsystem (for search and access of collection metadata); and provided to Interoperability Subsystem (for use in advertisements). ECS also uses collection metadata to update the system-wide data dictionary. Constructing collection metadata requires researching collection characteristics, choosing appropriate attributes, and assigning specific values.

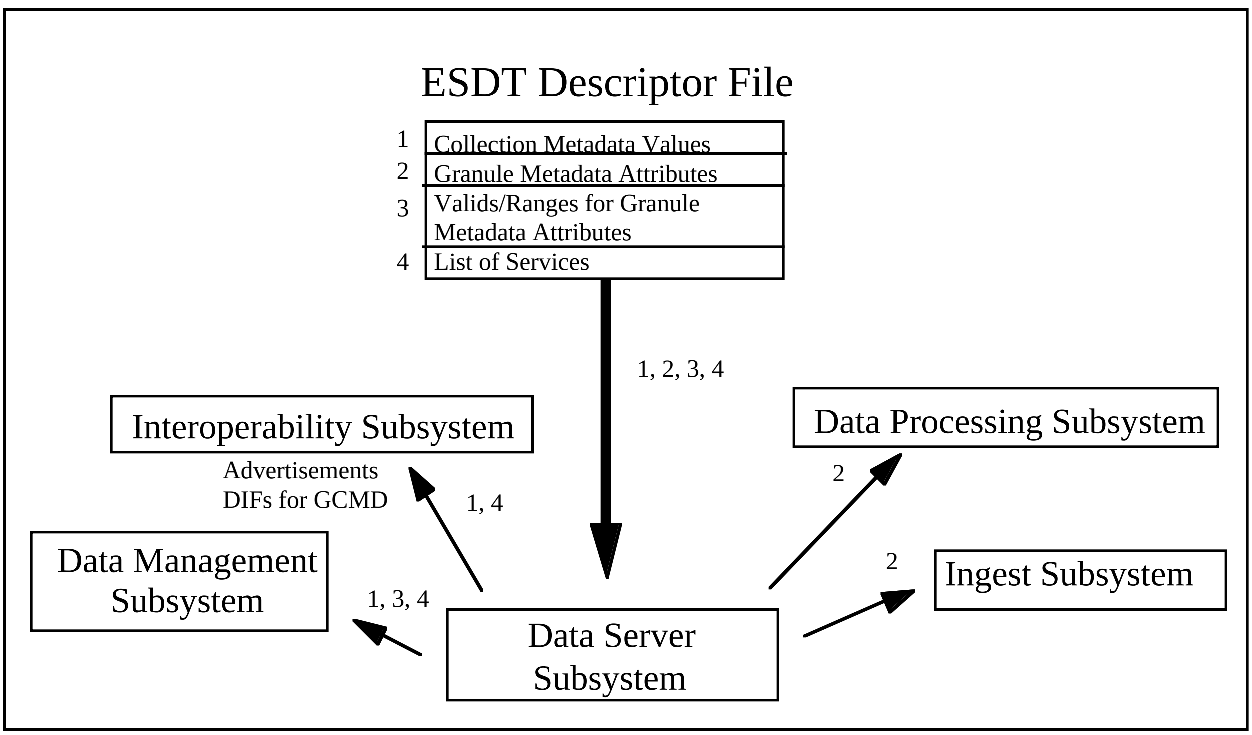

Figure 5-1. An ESDT Descriptor File is Parsed into Components Which are Used Throughout ECS

Granule metadata are stored in inventory tables within Science Data Server (for distribution with ordered data and also for search and access of granule metadata). Granule metadata attributes that describe each member of the collection must be identified in advance but specific values are mostly unknown at the time of constructing the ESDT descriptor file. Granule attributes (without values) are included in the ESDT descriptor so that Science Data Server can provide a template that is to be filled in during ingest processing (for external data delivered to ECS) or filled in during EOS instrument processing (for products produced within ECS). This template is called a Metadata Configuration File (MCF) and its use is further described in the two scenarios below: the first scenario for external data delivered to ECS and the second scenario for data produced within ECS. The third scenario describes how science users can use ECS metadata to search, access, and ultimately order science data of interest from the large, distributed EOSDIS data holdings.

Of note is the large difference in total number of attribute values between collection and granule metadata. A collection generated within ECS typically has ~100 collection attribute values but may easily contain 300,000 or more granule attribute values when summed across all granules in the collection. This large difference has significant implications for preparation and quality control of metadata. An error in a collection attribute or value can be corrected by manual edits to a single ESDT descriptor file; an error in a granule attribute or value will likely propagate through all members of the collection in the inventory tables. Contamination of the inventory can result in significant rework including the possibility of reprocessing. Data providers and producers should exercise special care when selecting and populating granule attributes that represent their data.

5.1 Science Data Ingest and Archival Scenario

There will be multiple sources of science data to be delivered to ECS for ingest and archive, such as EOS data (e.g. AM1 L0 data); ancillary data (e.g. NOAA data), Data Assimilation Office data; interdisciplinary science data; and EOSDIS V0 data (e.g. AVHRR Land Pathfinder data). Provision of ECS services, such as search and order, requires that values for granule metadata attributes must be extracted or derived during ingest processing and inserted into the Science Data Server DBMS inventory metadata tables. This scenario describes the process for extracting or deriving the needed values during ingest.

The scenario assumes that an ESDT descriptor file for the collection has already been inserted into the Science Data Server and Ingest is ready for external delivery of data granules belonging to the collection. The scenario begins with the arrival of the data and associated metadata files from an external source which is shown as Step (1) in Figure 5-2. Delivered data can be in HDF, HDF-EOS or binary format. Associated metadata are embedded in HDF and HDF-EOS; metadata for binary data may be found in the header file or in an additional ASCII file. Ingest notifies Science Data Server of the arrival of the data and in Step (2) the Science Data Server sends Ingest a Metadata Configuration File to be filled in with values for the granule metadata attributes. In Step (3) values are extracted or derived during Ingest Processing from the supplied data and metadata files according to the Metadata Location Tables which are stored by Ingest. The Metadata Location Tables are constructed from the Interface Control Document ICD)

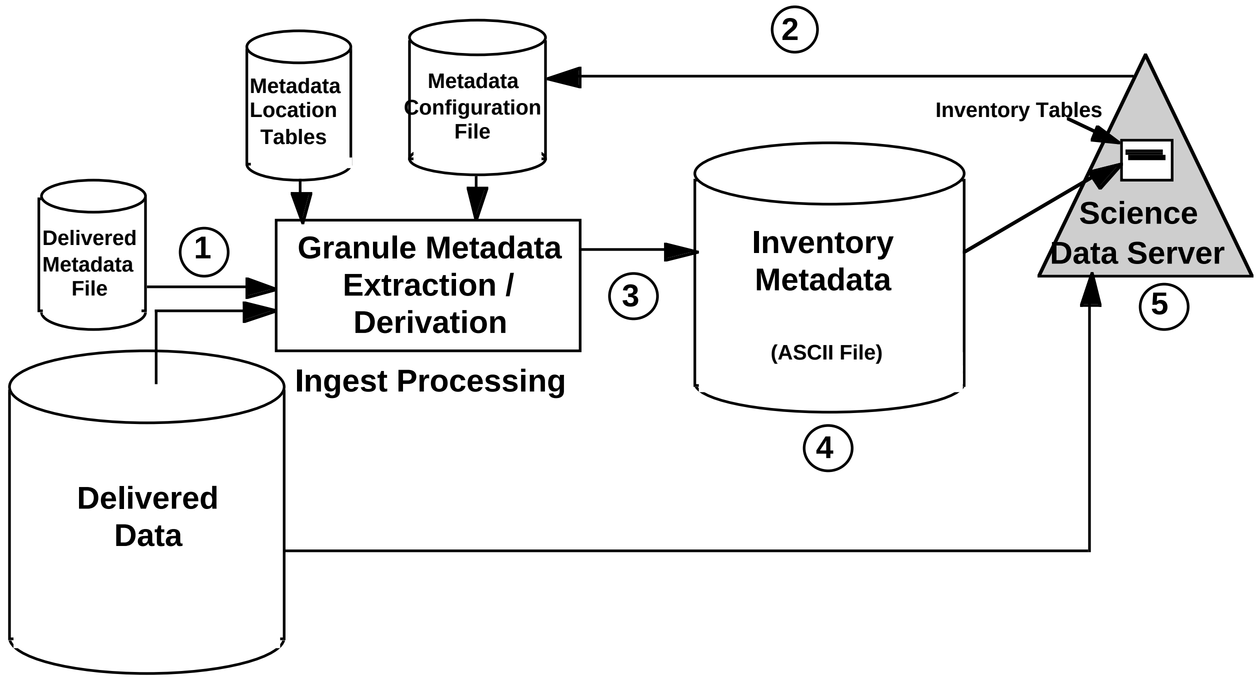

Figure 5-2. Science Data Ingest and Archival Scenario

negotiated between ECS and the Data Provider. In Step (4) the values for the granule metadata attributes are then written into the Metadata Configuration File template as inventory metadata (an ASCII file) and delivered to the Science Data Server. In Step (5) the Science Data Server first verifies that the inventory metadata can be inserted into the inventory tables, then the data granules are transferred to the Science Data Server for insertion into the archive.

5.2 Science Data Production and Archival Scenario

EOS instrument data that are to be processed within ECS (through 1999) include: AM1 instruments (ASTER, CERES, MISR, MODIS, MOPITT), a METEOR instrument (SAGE III), and Radar ALT instruments. Provision of ECS services requires that granule metadata must be generated during Science Data Processing and inserted into the Science Data Server inventory tables. This scenario describes the process for generating the granule metadata.

This scenario also assumes that an ESDT descriptor file for the collection has been inserted into the Science Data Server and that Science Data Processing is ready to produce granules that belong to the collection. This scenario begins with the arrival of inputs needed by Science Data Processing to produce new data granules for the collection which is shown as Step (1) in Figure 5-3. Inputs can be data and metadata from previous instrument processing, ancillary data and metadata, and other files needed for processing.

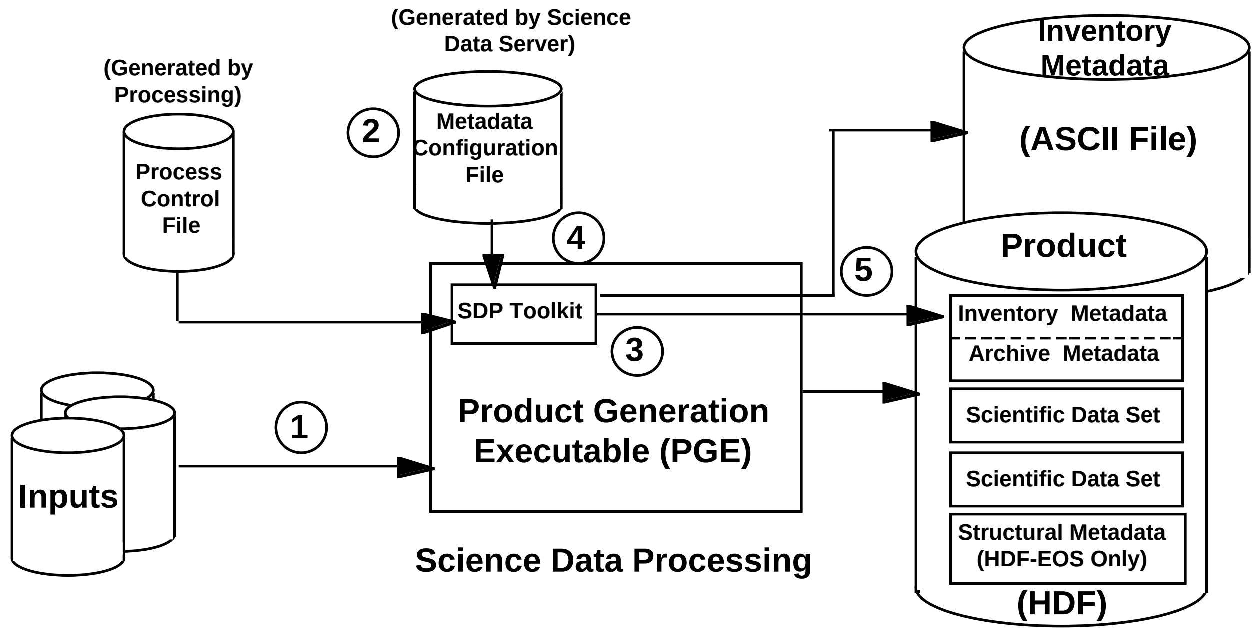

Figure 5-3. Science Data Production and Archival Scenario

Science Data Server Science notifies Data Processing of the arrival of the input data and when all the inputs are available, Science Data Processing then requests Science Data Server to return a Metadata Configuration File [Step (2)] that is to be filled in with values for the granule metadata attributes. In Step (3) Science Data Processing generates the new data granule (i.e., the "product") by running a Product Generation Executable (PGE) together with a Process Control File that defines the non-specific input and output file locations and other control parameters to the PGE. In Step (4) the PGE also produces the values for the granule metadata attributes by using the Science Data Processing (SDP) Toolkit to write values into two output metadata files. In Step (5) values are written into both the Metadata Configuration File template (inventory metadata) and into the HDF product (inventory and archive metadata). Archive metadata contain whatever information the data producer considers necessary to be in the HDF file, but to which system-wide services will not be applied. Insertion of the inventory metadata and the data granule into Science Data Server is the same as presented in the scenario in Section 5.1.

Note that there are also structural metadata automatically produced by a PGE for optional HDF-EOS Point, Swath, and Grid Structures. Structural metadata establish the relationships between data objects within an HDF-EOS file and support services on granules such as subsetting.

5.3 User Search, Access, Order, and Distribution Scenario

In general terms, data providers and data producers use ECS services to "push" data and metadata into the Science Data Server and other ECS subsystems so that distributed end users can "pull" the data and metadata for Earth science research. Sections 5.1 and 5.2 presented the data and metadata “push” scenarios; this section presents a user “pull” scenario showing user search, access, order, and distribution of EOSDIS data holdings.

The ECS Client provides the user’s view into the overall EOSDIS data repository. This is not a simple task, given the nature, size, and distribution of EOSDIS data holdings across multiple DAACs. However, by applying information management techniques tailored to EOSDIS metadata, ECS is able to provide users a very robust, interactive search environment. Users can efficiently assemble information about collections and connect to system-wide services such as search and access of metadata, and order and distribution of the data granules. Users are able to “drive down” into the data repository based upon their science research without a need to understand details about the directory and inventory tables or details about the data archives. There are many entry points into ECS through the ECS Client and the user has great flexibility to exercise and link various services during the information discovery process.

This “pull” scenario starts with the assumption of a new user who has little, if any, experience with EOSDIS. Other “pull” scenarios are certainly possible as discussed below after the presentation of the “new user” scenario. In Step (1) in Figure 5-4, the new user accesses the Interoperability Subsystem which appears to the user as the “Advertising Service” where general information about collections and services are advertised for all EOSDIS data holdings.

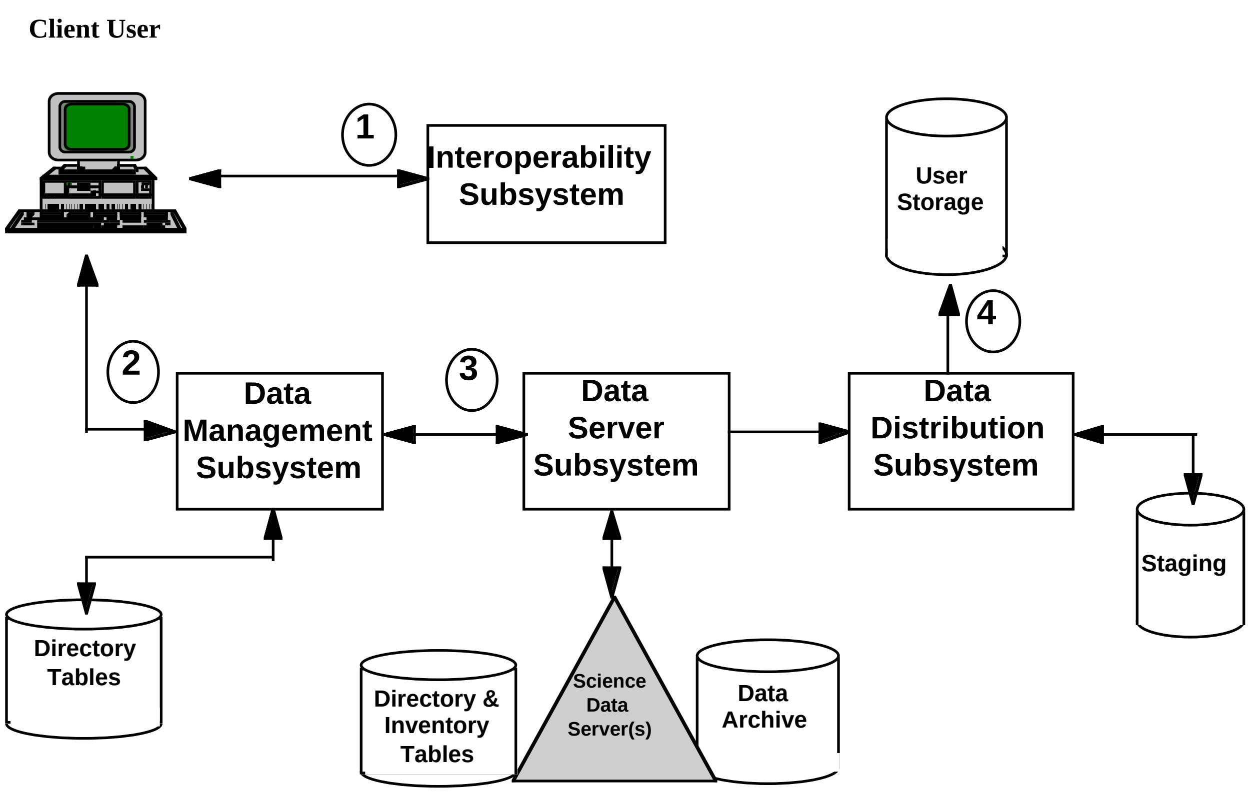

Figure 5-4. User Search, Access, Order, and Distribution Scenario

However, the level of visibility into the collections will be directly related to the metadata provided by the data providers and data producers. Metadata services may or may not be available depending on the provided metadata attributes. In addition to metadata services, the user will also have direct access to other documentation sources through online web access and links to key web sites.

Of importance in Steps (1) and (2) is the use of the Data Dictionary Service to provide useful metadata qualifiers, namely attribute valids and dependent valids. All collection metadata attributes have specific values or “valids” that aid the user in choosing collections of interest. In addition the Data Dictionary Service provides the Client with a view of which collection attributes are related to one another, i.e., the those that are ‘dependent’ upon one another. This enables the user to start with all available collections and through the use of valids and dependent valids to iteratively find the collections of interest based on specific science research objectives.

Step (2) represents the process of search and access of the collection and granule metadata. The ECS Client must pass through the Data Management Subsystem layer which acts like a “switchboard” mechanism to point the Client metadata query to the appropriate directory or inventory tables. Collection-level queries search the directory tables within the Data Management Subsystem. Granule-level queries search across distributed inventory tables at one or more DAACs. Hidden from the user is how the actual metadata query is compiled and exactly which path it takes as it is routed through the Data Management Subsystem.

In Step (3) the Data Server Subsystem accepts the query from the Data Management Subsystem; the Science Data Server component searches the appropriate inventory tables previously populated by the data providers and data producers. The result of the search is a list of granules that satisfy the search criteria. The user can select granules of interest and peruse the granule metadata. In addition, the user is shown what additional services exist for the data, including special services like Browse and Subsetting Services, and default services like Product Order Service. The user could request an additional data service to be performed or place a data order.

In Step (4) the user places a data order and the data are staged from the Science Data Server to the Data Distribution Subsystem. Depending on the user request, data are either sent via media to the user or staged for ftp pickup by the user.

The above scenario assumes a new user with little, if any, experience with EOSDIS. However, users will gain experience over time and others will have acquired considerable experience through use of EOSDIS Version 0. If a user knew the specific collection or granules of interest, the user could go directly to the metadata and data of interest. A query could be rapidly assembled by combining keywords together with spatial and temporal coverage. For example, the user may create a search for “MODIS sea surface temperature data over the Gulf of Mexico for January 2000.” The user could proceed with search and access of metadata, followed by order and distribution of the data as described in Steps (2) through (4) in Figure 5-4.

6. Responsibility for Preparation of EOSDIS Metadata

Metadata must be prepared both for data that will be delivered to ECS as well as for data that are to be produced within ECS. Collections in ECS will span all Earth science disciplines from land to oceans to atmosphere; interdisciplinary collections will span two or more of the Earth science disciplines. Due to the wide variety of data that will ultimately reside in ECS, varying responsibilities and skills will be required to prepare EOSDIS metadata. Specific pre-launch and post-launch responsibilities are in the process of being defined and this is one of the topics for the upcoming metadata workshop. Presented here is a general overview of participants and their approximate responsibilities. Note that most metadata preparation is a manual process utilizing the tools to be discussed in Section 7. The exception is granule metadata where values of metadata attributes are extracted either during ingest processing or produced during EOS instrument data processing as discussed in Section 5.

Because of the varying sources of data and metadata, not all metadata objects will be available or applicable to all data types. For example, no developed algorithm package will be delivered with ancillary data so no delivered algorithm package metadata will be resident in ECS for ancillary data. Shown in the Tables 6-1 to 6-4 are the varying responsibilities for metadata preparation beginning with granule and browse metadata for data to be delivered to ECS. Metadata are expected only for those metadata objects shown in **bold** in Table 6-1.

Table 6-1. Responsibility for Preparation of Granule and Browse Metadata for Data to be Delivered to ECS (1 of 2)

Data Type	Metadata Object	Responsibility for Preparation*	Metadata Delivery Format	Source of ECS Metadata
AM1 L0 data - ASTER - CERES - MISR - MODIS - MOPITT	**Granule** Browse	a, b, c not applicable	ICD for ECS-to-EDOS not applicable	Extract during ECS ingest based on ICD
SAGE III L0 data	**Granule** Browse	a, b, c not applicable	ICD for ECS-to-SAGE III not applicable	Extract during ECS ingest based on ICD
Radat ALT L0 data	**Granule** Browse	a, b, c not applicable	ICD for ECS-to-Radar ALT not applicable	Extract during ECS ingest based on ICD
Landsat 7 L0R data	**Granule** **Browse**	b, c, d b, c, d	ICD for ECS-to-Landsat 7 ICD for ECS-to-Landsat 7	Extract during ECS ingest based on ICD

* a = EOS instrument team, b = ECS staff, c = DAAC staff, d = data provider

Table 6-1. Responsibility for Preparation of Granule and Browse Metadata for Data to be Delivered to ECS (2 of 2)

Data Type	Metadata Object	Responsibility for Preparation*	Metadata Delivery Format	Source of ECS Metadata
SeaWinds data	**Granule** **Browse**	b, c, d b, c, d	ICD for ECS-to-SeaWinds ICD for ECS-to-SeaWinds	Extract during ECS ingest based on ICD
Ancillary data	**Granule** Browse	b, c not applicable	Native not applicable	Derive from native data and metadata during ECS ingest
Data Assimilation Office (DAO) data	**Granule** **Browse**	b, c, e b, c, e	ICD for ECS-to-DAO ICD for ECS-to-DAO	Extract during ECS ingest based on ICD
Interdisciplinary science (IDS) data	**Granule** **Browse**	c, f c, f	ICD for ECS-to-SCF ICD for ECS-to-SCF	Extract during ECS ingest based on ICD
EOSDIS V0 data	**Granule** **Browse**	c, g c, g	Section J, SDP Toolkit Section J, SDP Toolkit	Deliver metadata directly to Science Data Server

* b = ECS staff, c = DAAC staff, d = data provider, e = DAO, f = IDS, g = V0 data migration team

Shown in Table 6-2 is the responsibility for preparation of granule and browse metadata for EOS instrument data to be processed within ECS. Granule and browse metadata are expected for all EOS instrument data to be processed within ECS.

Table 6-2. Responsibility for Preparation of Granule and Browse Metadata for EOS Instrument Data to be Processed within ECS

Data Type	Metadata Object	Responsibility for Preparation*	Metadata Delivery Format	Source of ECS Metadata
AM-1 instruments - ASTER - CERES - MISR - MODIS - MOPITT	**Granule** **Browse**	a, b, c a, b, c	Section J, SDP Toolkit Section J, SDP Toolkit	Produced by PGEs Produced by PGEs
METEOR instrument - SAGE III	**Granule** **Browse**	a, b, c a, b, c	Section J, SDP Toolkit Section J, SDP Toolkit	Produced by PGEs Produced by PGEs
Radar ALT instruments	**Granule** **Browse**	a, b, c a, b, c	Section J, SDP Toolkit Section J, SDP Toolkit	Produced by PGEs Produced by PGEs

* a = EOS instrument team, b = ECS staff, c = DAAC staff

Shown in Table 6-3 is the responsibility for preparation of collection, delivered algorithm package, and document metadata for data to be delivered to ECS**.** Metadata are expected only for those metadata objects shown in **bold** in Table 6-3.

Table 6-3. Responsibility for Preparation of Collection, Delivered Algorithm Package, and Document Metadata for Data to be Delivered to ECS

Data Type	Metadata Object	Responsibility for Preparation*	Metadata Delivery Format
AM1 L0 data	**Collection** Delivered Algorithm Package Documents	a, b, c not applicable not applicable	MetaDataWorks not applicable not applicable
SAGE III L0 data	**Collection** Delivered Algorithm Package Documents	a, b, c not applicable not applicable	MetaDataWorks not applicable not applicable
Radar ALT L0 data	**Collection** Delivered Algorithm Package Documents	a, b, c not applicable not applicable	MetaDataWorks not applicable MetaDataWorks
Landsat 7 L0R data	**Collection** Delivered Algorithm Package **Documents**	b, c, d not applicable c, d	MetaDataWorks not applicable MetaDataWorks
SeaWinds data	**Collection** Delivered Algorithm Package **Documents**	b, c, d not applicable c, d	MetaDataWorks not applicable MetaDataWorks
Ancillary data	**Collection** Delivered Algorithm Package Documents	b, c not applicable not applicable	MetaDataWorks not applicable not applicable
Data Assimilation Office (DAO) data	**Collection** Delivered Algorithm Package **Documents**	b, c, e not applicable c, e	MetaDataWorks not applicable MetaDataWorks
Interdisciplinary science (IDS) data	**Collection** Delivered Algorithm Package **Documents**	c, f not applicable c, f	MetaDataWorks not applicable MetaDataWorks
EOSDIS V0 data	**Collection** Delivered Algorithm Package **Documents**	c, g not applicable c, g	MetaDataWorks not applicable MetaDataWorks

* a = EOS instrument team, b = ECS staff, c = DAAC staff, d = data provider, e = DAO, f = IDS, g = V0 data migration team

Shown in Table 6-4 is the responsibility for preparation of collection, delivered algorithm package, and document metadata for EOS instrument data to be processed within ECS. Collection, delivered algorithm package, and document metadata are expected for all EOS instrument data to be processed within ECS.

Table 6-4. Responsibility for Preparation of Collection, Delivered Algorithm Package, and Document Metadata for EOS Instrument Data to be Processed within ECS

Data Type	Metadata Object	Responsibility for Preparation*	Metadata Delivery Format
AM-1 instruments - ASTER - CERES - MISR - MODIS - MOPITT	**Collection** **Delivered Algorithm Package** **Documents**	a, b, c a, c a, c	MetaDataWorks MetaDataWorks MetaDataWorks
METEOR instrument - SAGE III	**Collection** **Delivered Algorithm Package** **Documents**	a, b c a, c a, c	MetaDataWorks MetaDataWorks MetaDataWorks
Radar ALT instruments	**Collection** **Delivered Algorithm Package** **Documents**	a, b, c a, c a, c	MetaDataWorks MetaDataWorks MetaDataWorks

* a = EOS instrument team, b = ECS staff, c = DAAC staff

7. Tools and Processes for Preparing EOSDIS Metadata

During early ECS development, the EOS instrument teams were unfortunately subjected to a heavy dose of engineering documents and activities as they attempted to prepare their first set of metadata attributes and values. To help alleviate these early problems, ECS has expanded its metadata consulting staff so that each EOS instrument team has its own ECS metadata consultant. Also, tools are being built to streamline metadata entry and thereby reduce the need to master the ECS-ready Object Description Language (ODL). Finally, a clear schedule for metadata population has been developed and communicated to the EOS instrument teams, and we are implementing a standard process for working together to jointly meet that schedule.

7.1 Tools for Preparing EOSDIS Metadata

ECS is developing several key tools to aid the EOS instrument teams and other persons responsible for preparing metadata. The overall goals in building these tools are:

a. Aid in verifying the accuracy of metadata through as much automated validation as possible.

b. Improve online reference materials.

c. Reduce efforts of the EOS instrument teams in providing metadata by moving the computer science, or implementation details, to ECS and DAAC staff.

Tools development will be an ongoing process; the initial tool set will support immediate metadata population and will lead ultimately to a capability for automated generation of descriptor files for new or updated ESDTs. These initial tools and their functions are:

The ECS Data Dictionary: This user-friendly online reference displays information about the ECS Data Model such as grouping, data types, and attribute definitions. Completed in February 1997, the ECS Data Dictionary is already proving useful to ECS staff supporting the AM1 instrument teams.

The Product Specific Attribute Registry: This online registry is an outgrowth of the improved Product Specific Attribute process. This Registry can review existing Product Specific Attributes or add/or update new ones. In addition, the database underlying this product is used to produce timely reports to EOS instrument teams on particular Product Specific Attributes, as well as all attributes submitted. The Product Specific Attribute Registry was completed in early March, 1997.

The Shortname Registry: This registry captures reference names for collections, platforms, campaigns, etc. The purpose of the registry is to ensure uniqueness and commonality of definition.

MetaDataWorks: An outgrowth of an early prototype for metadata population called M_POP, MetaDataWorks is a Web-based metadata entry system. Extracting the attributes and valid values from the online data dictionary, MetaDataWorks is a flexible system for building Earth Science Data Type descriptors. It uses business rules (such as rules for setting links between attributes) as well as a set of defaults defined by AM1 instrument teams, to quickly build ECS-ready metadata. MetaDataWorks is in beta testing in the month of March 1997.

Metadata Configuration File Builder. EOS instrument teams need no longer hand-write Metadata Configuration Files (see Section 5) for each product generation executable (PGE, see Section 5). As part of metadata entry process, each Metadata Configuration File is automatically built from the metadata entered into MetaDataWorks. The first version of this tool is scheduled for completion in late March 1997. The second version is scheduled for completion in late May 1997.

Additional tools used by ECS to validate prepared metadata include an ODL parser that reads metadata in ODL structure, validates the syntax and structure, and indicates any errors. Another tool uses a local version of the Science Data Server to validate metadata and services. In the first phase of this validation, the ESDT is entered into the local version of the Science Data Server. Next, an appropriate simulated data granule is inserted into the Science Data Server and services are exercised to validate metadata, simulated data, and available services.

7.2 Processes for Preparing EOSDIS Metadata

Figure 7-1 shows the steps involved in pre-launch preparation and validation of an ESDT descriptor file for EOS instrument processing. Initially this is a joint process between the EOS instrument teams and ECS staff with later involvement of DAAC staff. Responsibility for preparing ESDT descriptors is a topic for discussion at the Pre-Launch Metadata Workshop. As discussed in Section 4.4, some data types, such as interim files, require minimum metadata; whereas, deliverable EOS granules require a richer set of metadata to support more advanced services. After delivery of ECS to the DAACs, all tools and processes will be transferred to the DAACs for their use in future EOS operations.

ESDT Descriptor File Preparation and Validation for EOS Instrument Processing

Step	EOS Instrument Team	ECS Staff
• Step 1. Needs Analysis	Joint	
• Step 2. Finalize Schedule	Joint	
• Step 3. Prepare Supporting Information	Joint	
• Step 4. Review Product Specific Attributes	Joint	
• Step 5. Compile and Organize Metadata		✔
• Step 7. Jointly Validate the Metadata	Joint	
• Step 8. Receive the Metadata Config File		✔
• Step 9. Build Product Services		✔
• Step 10. Validate in ECS		✔
• Step 11. Use Metadata in Algorithms	Joint	
• Step 12. Update Metadata	Joint	

Figure 7-1. The Steps Involved in Preparing and Validating an ESDT Descriptor File for EOS Instrument Processing

This page intentionally left blank.

8. Maintenance and Updates of EOSDIS Metadata

The ECS Science Data Server retains all inserted metadata information and forwards metadata information to other ECS subsystems as previously discussed in Section 5. This process updates metadata information throughout ECS as new data are entered into ECS. The methods for changing metadata vary depending on the type of metadata: collection or granule.

Collection metadata changes: When the value of an existing collection metadata attribute needs to changed, the change is submitted to the DAAC change control board according to published DAAC policies. After approval, a new ESDT descriptor file is created and inserted by DAAC operations as a new version.

Granule metadata changes: Updates of the QA flags are performed using the QA support tools by operations staff at the DAACs as well as by science instrument teams at their remote locations.

Maintenance of both collection and granule metadata can be performed using a commercial-off-the-shelf GUI interface that is provided with the DBMS. This software-level interface allows operations staff at the DAACs to accomplish maintenance of the directory and inventory tables, as necessary.

For each data granule ordered from ECS, the Science Data Server forwards the filename to Data Distribution Service for distribution to the end user. These are the files that reside in the archive. Additionally, the Science Data Server will stream out the granule's current collection and granule metadata values to a file, in ODL format. This file is added to the list of files that Data Distribution Service is asked to distribute. Data Distribution Service will then ask Storage Management Service to retrieve the archived files, and stage the metadata files for distribution.

When a collection is subsetted, new metadata describing the subsetted collection will be prepared for distribution to the requester. Subsetted files are not stored in the archive, nor is the new metadata stored in the inventory tables. The Science Data Server refers Data Distribution Service to the subsetted files in working storage for distribution to the requester.

The Science Data Server does not embed metadata in non-HDF files. All of the data that are non-HDF will be sent with an accompanying metadata file formatted using ODL. A complete set of the metadata will be delivered, reflecting the current state of the data granule, not just metadata changes or deltas. At this time there is no general agreement within the customer community regarding rules for embedding "current state of the data granule" metadata within HDF-EOS granule files. ESDIS, the user community, and ECS are currently working together to define the rules. Note that the end user can inspect granule and collection metadata before placing an order via the Client Service.

This page intentionally left blank.